МЕТОДЫ ОБРАБОТКИ СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И СИГНАЛОВ
(Вычислительные системы)
1989 год Выпуск 132

УДК 519.237.5

ПРАКТИЧЕСКИЕ АЛГОРИТМЫ ОПТИМИЗАЦИИ И АППРОКСИМАЦИИ ФУНКЦИЙ

В.Г. Устюжанинов

Введение

В работе обобщается опыт шестилетней эксплуатации ППП "Синтез" [1] в решении задач оптимизации и аппроксимации функций в условиях, когда последние заданы операторным способом.

Два обстоятельства определили специфику нашего подхода к решению этих задач.

Во-первых, мы не располагали предварительной информацией о задаче, необходимой для вычисления оценок качества ее решения в рамках теории регрессионного анализа [2].

Ввиду того, что число точек в обучающей выборке обычно меньше числа регрессоров, используемых для построения регрессии, нам пришлось по-новому взглянуть на постановку задачи, способ ее решения и на то, что, собственно, считать результатом решения задачи. Отсюда и наше отступление от традиционных представлений математиков.

В частности, такой подход характеризуется следующими особенностями. Формальная запись задачи рассматривается лишь как цель, к которой надлежит стремиться, а не как математический атрибут, используемый в теоремах. Алгоритмы решения задачи носят эвристический характер. Не делается попытки доказать оптимальность решения в том или ином смысле. Достаточно мнения экспериментатора о том, что полученное решение для него приемлемо.

Во-вторых, немаловажное значение для выбора нашего подхода имела значительная стоимость отдельного эксперимента. Это привело к необходимости минимизации числа экспериментальных точек, требуемых для решения задачи,и, как следствие, к использованию последовательных алгоритмов, в которых этапы экспериментальной работы перемежаются с этапами анализа экспериментальных данных и этапами планирования эксперимента.

Может показаться, что отсутствие теоретических оценок качества решения, даваемого алгоритмом, неизбежно отрицательно скажется на его эксплуатации.В общем случае это опасение справедливо. Однако практика эксплуатации пакета показала,что представляемых пользователю средств достаточно, чтобы экспериментаторы нащупывали новые эффекты и даже открывали новые для себя направления работ.

§2. Постановка задачи

Пусть Q - компактная область евклидова пространства R^k, элементами которого являются векторы x ,в дальнейшем именуемые точками. Задан класс функций Ψ с элементами: $\psi\colon Q \to R^1$, где ψ непрерывно на Q . Имеется "черный ящик" с одним входом и одним выходом. Если на вход "черного ящика" подать точку $x \in Q$, то выход примет значение в соответствии с уравнением

$$y(x) = \sigma(x) + \xi , \tag{1}$$

в котором величина $\xi \in R^1$ есть погрешность, вносимая "черным ящиком", а в отношении σ известно, что $\sigma \in \Psi$.

Задача оптимизации. Пусть $\Phi = \max\limits_{x \in Q} \psi(x)$. Фиксируется число $\varepsilon > 0$. С каждой $\psi \in \Psi$ ассоциируется множество точек $M(\psi) = \{ x / \Phi - \psi(x) \le \varepsilon,\ x \in Q \}$.

Требуется указать точку из $M(\sigma)$.

Задача аппроксимации. Фиксируется базисная система функций $\Phi=(\varphi_1,\ldots,\varphi_n)$, где $\varphi_i: Q\to R^1$, $i=1,\ldots,n$, и непрерывны на Q. На функции $\varphi_i \in \Phi$ натянуто n-мерное линейное пространство F обобщенных многочленов

$$f(x) = \sum_{i=1}^{n} a_i\varphi_i(x) \tag{2}$$

с коэффициентами $a_i \in R^1$.

В дальнейшем обобщенные многочлены именуются просто многочленами.

На множестве пар $(\psi,f)\in \Psi\times F$ определяется функция расстояния

$$\rho(\psi,f) = \max_{x\in Q} |\psi(x)-f(x)|.$$

Фиксируется число $\varepsilon > 0$. С каждой $\psi\in \Psi$ ассоциируется множество многочленов $M(\psi)=\{f/f\in F,\ \rho(\psi,f)\le \varepsilon\}$.

Классы Ψ, F и число ε выбираются так, чтобы для любой $\psi\in \Psi$ множество $M(\psi) \neq \emptyset$.

Требуется указать многочлен из $M(\psi)$.

§3. Оптмизация

Процесс решения задачи состоит из ряда следующих друг за другом этапов. Первый из них проходит пять стадий: выбор начального плана эксперимента; проведение эксперимента в точках плана; генерация семейства ε-тупиковых многочленов V_ε; выбор множества новых экспериментальных точек X; проведение эксперимента в точках X. Остальные этапы имеют три стадии - это генерация семейства V_ε, выбор множества точек X и проведение эксперимента в точках множества X. В практической деятельности для выбора начального плана мы использовали руководство [3].

Для чего нужны ε-тупиковые многочлены? ε-тупиковый многочлен - это функция $f \in F$,аппроксимирующая функцию σ с качеством ε на точках плана. По отношению к каждому из ε-тупиковых многочленов,входящих в V_ε , решается задача оптимизации на области Q и находятся их точки максимума. Из них составляется множество новых экспериментальных точек X . В этих точках проводится эксперимент. Если среди ε-тупиковых многочленов, входящих в V_ε , найдется такой, который хорошо аппроксимирует σ не только на точках плана, но и во всей области Q , то среди точек $x \in X$ окажется и точка x^* из $M(\sigma)$, при проведении эксперимента в которой будет зафиксирован приемлемый для экспериментатора результат.

Переходим к подробному описанию произвольного r-го $(r>1)$ этапа. Пусть к моменту его исполнения проведены измерения функции σ в точках $x_1, \dots, x_m$ и получены значения $y_j = y(x_j)$ согласно уравнению (1). Качество аппроксимации функции σ многочленом $f \in F$ определим как

$$\varepsilon(f) = \frac{1}{m} \sum_{j=1}^{m} (y_j - f(x_j))^2 .$$

Функцию f поставим в соответствие вектору коэффициентов $a(f) = (a_1, \dots, a_n)$, участвующих в разложении (2) многочлена f по системе базисных функций Φ .Образуем множество $I(f)$, состоящее из номеров ненулевых компонент вектора $a(f)$. Зафиксируем $\varepsilon > 0$.

ОПРЕДЕЛЕНИЕ. Многочлен f назовем ε-тупиковым, если

1) $\varepsilon(f) \le \varepsilon$;

2) для всех $g \in F$ с $I(g) = I(f)$ величина $\varepsilon(g) \ge \varepsilon(f)$;

3) для всех $g \in F$ с $I(g) \subset I(f)$ величина $\varepsilon(g) > \varepsilon$.

Число $s(f) = \|I(f)\|$ назовем сложностью многочлена f.

В практическом выборе величины ε мы руководствовались следующим правилом. Оценивалась дисперсия D случайной величины ξ в (1), и величина ε выбиралась примерно равной D.

Количество ε-тупиковых многочленов конечно. Упорядочим их по возрастанию сложности. Отсечем в этой последовательности первые q штук многочленов и образуем из них класс W_ε ε-тупиковых многочленов минимальной сложности. На практике число таких многочленов доходило до 1000 штук, а число входящих в них базисных функций - до 200. Отбор многочленов в класс W_ε осуществлялся по алгоритму, описанному в [4].

Учтем требования, предъявляемые к многочленам $f \in W_\varepsilon$ со стороны экспериментатора. Ему могут быть известны детали поведения функции σ на области Q. Например, неотрицательность σ или ее частных производных на Q. Это именно те требования, которые определяют класс Ψ. Требования, предъявляемые к σ, переносятся на многочлены $f \in W_\varepsilon$. Те из них, которые этим требованиям не удовлетворяют, удаляются из W_ε. Получаем класс многочленов V_ε.

Выберем в Q конечную сетку точек G. Каждому многочлену $f \in V_\varepsilon$ поставим в соответствие единственную точку z, доставляющую решение задачи $f(x) \to \max\limits_{x \in G}$. Разобьем класс V_ε на подклассы $V_\varepsilon(z)$, относя к $V_\varepsilon(z)$ те многочлены из V_ε, которым поставлена в соответствие точка z. Назовем весом точки z число $p(z) = \|V_\varepsilon(z)\|$. Упорядочим точки $z \in G$ по уменьшению их веса. Получим последовательность $z_1, z_2, \ldots$ и т.д.

На практике постоянно приходилось работать с сетками, содержащими около 10 тысяч точек. Неожиданным оказалось то, что число точек с ненулевым весом было обычно в пределах двух десятков. Причем около 70-90% суммарного веса сосредоточивалось на первых четырех точках. Эта четверка обладала важными свой -

ствами для проведения в них эксперимента. Если ожидание максимального значения σ хотя бы в одной из точек z_1, z_2, z_3, z_4 подтверждалось, то экспериментатор получал приемлемый для него результат и мог прекратить дальнейший поиск. Если же ожидание не подтверждалось, то, введя эти точки в обучение, мы добивались отсеивания 70-90% многочленов из класса V_ε. Это должно было сократить число этапов работы алгоритма и как следствие сэкономить труд экспериментатора.

Для эксперимента предлагались не сами точки последовательности $z_1, z_2, \ldots$, а их модификации. Пусть точке z_i соответствует непустой класс $V_\varepsilon(z_i)$. Вычислим вектор $b = (b_1, \ldots, b_n)$ по формуле

$$b = \frac{1}{\|V_\varepsilon(z_i)\|} \sum_{f \in V_\varepsilon(z_i)} a(f).$$

Введем функцию

$$g_i(x) = \sum_{j=1}^{n} b_j \varphi_j(x).$$

Назовем $g_i(x)$ моделью-представителем класса $V_\varepsilon(z_i)$. Найдем точку максимума $g_i(x)$ на Q. Обозначим ее z_i^*. Присвоим z_i^* вес точки z_i. Упорядочим точки z_i^* ($i = 1, 2, \ldots$) по их весу. Получим последовательность $z_{\nu_1}^*$, $z_{\nu_2}^*, \ldots$ и т.д. Отсечем первые t точек в этой последовательности и составим из них множество новых экспериментальных точек X. На практике t обычно было 4-5.

По отношению к конкретной точке z_i^* возможны следующие ситуации.

1. Точка $z_i^* \in Q$, но экспериментатор утверждает, что провести эксперимент в z_i^* невозможно, поэтому следует ввести дополнительные ранее не очевидные ограничения на Q.

2. Экспериментатор утверждает, что получить в точке z_i^* приемлемый для него результат невозможно. Это противоречит его научному опыту. Но в случае получения приемлемого экспериментального результата в z_i^* происходит прорыв в новую научную область.

3. В точке z_i^* получен приемлемый результат. Решение задачи прекращается.

Наконец, отметим, еще одно важное обстоятельство. Если в процессе построения множества V_ε оно оказалось пустым, то это означает, что принятая система базисных функций не обеспечивает требуемого качества аппроксимации функции σ. Необходима смена системы базисных функций.

§4. Аппроксимация

По методу, описанному в §3, генерировался класс V_ε ε-тупиковых многочленов. Далее проводилась квантификация значений функций $f \in V_\varepsilon$. Для этого фиксировалась конечная сетка точек $G \subset Q$. Для каждой точки $x \in G$ находились числа $\bar{y}(x) = \min\limits_{f \in V_\varepsilon} f(x)$ и $\bar{\bar{y}}(x) = \max\limits_{f \in V_\varepsilon} f(x)$. Фиксировалось целое число N (на практике $N = 10$). Квант значений w функции f в точке x определялся по формуле $w(x) = \max\{d, (\bar{\bar{y}}(x) - \bar{y}(x))/N\}$, где d выбиралось из интервала $[\varepsilon, 2\varepsilon]$. Интервал $[\bar{y}(x), \bar{\bar{y}}(x)]$ разбивался на систему интервалов $L_i(x) = \{y / \bar{y}(x) + w(x)(i-1) \le y < \bar{y}(x) + w(x)i\}$, $i = 1, \dots, N-1$, и интервал $L_N = \{y / \bar{y}(x) + w(x)(N-1) \le y \le \bar{\bar{y}}\}$. Каждому многочлену $f \in V_\varepsilon$ ставилась в соответствие функция $g_f : G \to \{1, 2, \dots, N\}$. Значения g_f вычислялись по правилу: $g_f(x) = i$, если $f(x) \in L_i(x)$. Строилась таблица τ, строкам которой соответствовали многочлены $f \in V_\varepsilon$, а столбцам - точки $x \in G$. На пересечении строки f и столбца x располагался элемент $g_f(x)$. Размеры таблицы доходили до 10^3 строк и $5 \cdot 10^3$ столбцов. Затем на таблице τ разыскивался так называемый минимальный тест. Что это такое?

Набор столбцов $x_{\nu_1},\dots,x_{\nu_s}$ таблицы τ назовем тестом, если в образованной этими столбцами подтаблице все строки попарно различны. Число s - длина теста. Тест назовем минимальным, если его длина минимальна среди всех тестов таблицы τ.

Поиск минимального теста осуществлялся с помощью градиентной процедуры из [5] путем предварительного сведения задачи поиска минимального теста к задаче поиска минимального покрытия бинарной таблицы.

Тест обладает важным свойством. Пусть точки теста $x_{\nu_1},\dots,x_{\nu_s}$ были поданы на вход "черного ящика" и в результате можно было наблюдать значения $y_1,\dots,y_s$. Допустим, что $y_i \in L_{\eta_i}(x_{\nu_i})$ для $i = 1,\dots,s$. Рассмотрим произвольный многочлен $f \in V_\varepsilon$. Будем говорить, что серия экспериментов в точках $x_{\nu_1},\dots,x_{\nu_s}$ подтвердила многочлен f, если для всех $i = 1,\dots,s$ значения $f(x_{\nu_i}) \subset L_{\eta_i}(x_{\nu_i})$. Понятно, что в этом случае $g_f(x_{\nu_i}) = \eta_i$ для всех $i = 1,\dots,s$. Иначе говорим, что многочлен f отвергается экспериментами в точках $x_{\nu_1},\dots,x_{\nu_s}$. Из способа получения теста следует, что проведение в его точках серии экспериментов либо отвергает все многочлены из класса V_ε, либо подтверждает не более одного из них, отвергая все остальные.

Таким образом, тест является мощным средством в идентификации аппроксимирующего многочлена. Использование его для проведения эксперимента должно привести к уменьшению числа этапов алгоритма, к уменьшению объема экспериментальной работы. По той же причине нами использовались не произвольные тесты, а минимальные.

Есть возможность еще уменьшить число точек, рекомендуемых для эксперимента, если перейти от тестов к использованию частичных тестов. Дадим их определение.

Набор столбцов $x_{\nu_1}, \dots, x_{\nu_s}$ таблицы τ назовем частичным тестом, если в образованной этими столбцами подтаблице имеются пары одинаковых строк.

Зафиксируем набор $x_{\nu_1}, \dots, x_{\nu_s}$. Выделим в τ таблицу τ', состоящую из столбцов $x_{\nu_1}, \dots, x_{\nu_s}$. Пусть $q = \|V_\varepsilon\|$. Тогда число строк в τ' есть q. Число всевозможных пар строк будет C_q^2. Подсчитаем число пар неодинаковых строк в таблице τ'. Обозначим его $S(x_{\nu_1}, \dots, x_{\nu_s})$ и назовем различием. Зафиксируем s и найдем частичный тест $x^*_{\nu_1}, \dots, x^*_{\nu_s}$, на котором достигается максимум $S(x_{\nu_1}, \dots, x_{\nu_s})$ на всевозможных наборах столбцов $x_{\nu_1}, \dots, x_{\nu_s}$. Назовем $x^*_{\nu_1}, \dots, x^*_{\nu_s}$ оптимальным частичным тестом.

На практике для поиска оптимального частичного теста нами использовалась градиентная процедура из [5].

Применение частичного теста дает резкое сокращение числа экспериментальных точек, требуемых для идентификации регрессии σ. Так, в одной из задач при поиске частичного теста для таблицы τ из 996 строк и 4000 столбцов были найдены частичные тесты со следующими различиями: тест из одной точки - различие 434419; тест из двух точек - различие 484459; тест из трех точек - различие 492964; тест из четырех точек - различие 494770. Заметим, что $C_{996}^2 = 496008$. Понятно, что экспериментирование с тестом даже из двух точек приведет к резкому сокращению числа многочленов из V_ε, претендующих на роль аппроксимирующих функцию σ.

Пусть $\{x_{\nu_1},\ldots,x_{\nu_s}\}\subseteq\{x_{\mu_1},\ldots,x_{\mu_l}\}$. Тогда $S(x_{\nu_1},\ldots,x_{\nu_s})\leq S(x_{\mu_1},\ldots,x_{\mu_l})$. Отсюда следует, что максимальное различие достигается на тесте, состоящем из всех столбцов таблицы τ. Пусть τ имеет q строк и n столбцов. На практике могут встретиться такие таблицы, в которых $S(x_1,\ldots,x_n)\leq C_q^2$. В этом случае использование частичных тестов становится неизбежным.

Опишем ситуации, возникающие в процессе работы алгоритма, и те решения, которые мы в этих ситуациях принимали.

1. Пусть множество многочленов V_ε пусто. Это означает, что семейства базисных функций Φ недостаточно для качественной аппроксимации функции σ. Необходим переход к другому семейству базисных функций.

2. Пусть на очередном этапе проведены эксперименты в точках теста $x_{\nu_1},\ldots,x_{\nu_s}$.

2.1. Оказалось, что все многочлены, входящие в V_ε, этим тестом отвергнуты. Решение - надо точки теста включить в обучающую выборку и, используя последнюю, построить новое множество многочленов V_ε.

2.2. Оказалось, что некоторая часть многочленов $\mathbb{T}$ из V_ε удовлетворяет точкам теста $x_{\nu_1},\ldots,x_{\nu_s}$. Возникает вопрос, следует ли продолжать поиск аппроксимирующего многочлена или закончить процесс, назвав любой многочлен из $\mathbb{T}$ в качестве результата работы алгоритма. Это вопрос о критерии останова.

Надо сказать, что ответ на него на практике лежит вне компетенции математика. Это связано с тем, что аппроксимирующая функция нужна не сама по себе, а как инструмент достижения прикладной цели. Например, с помощью аппроксимирующей функции могут прогнозироваться свойства прикладных объектов. В зависимости от того, удачен ли этот прогноз или нет принимается ре -

шение об останове процесса поиска аппроксимирующего многочлена или его продолжении.

Литература

1. УСТЮЖАНИНОВ В.Г., ШЕМЯКИНА Е.Н. Пакет прикладных программ СИНТЕЗ для построения и анализа моделей прогноза //Методы анализа данных.-Новосибирск,1985. -Вып.111: Вычислительные системы. -С. 77-89.

2. ДЕМИДЕНКО Е.З. Линейная и нелинейная регрессия. - М.: Финансы и статистика, 1981. - 302 с.

3. ХАРТМАН К. и др. Планирование эксперимента в исследовании технологических процессов. - М.: Мир, 1977.

4. УСТЮЖАНИНОВ В.Г., ШЕМЯКИНА Е.Н. Алгоритм выбора линейной регрессии минимальной сложности //Анализ данных в экспертных системах. - Новосибирск, 1986. - Вып. 117: Вычислительные системы. - С. 78-82.

5. ЧЕГИС И.А., ЯБЛОНСКИЙ С.В. Логические способы контроля работы электрических схем //Труды математического института им. В.А.Стеклова. 1958. Т. 51. - С. 270-360.

Поступила в ред.-изд.отд.
19 июня 1989 года